

## チームワーク形成を目的としたPBL型教育の効果測定：研修評価アプローチによる検討

著者	安田 節之, 梅崎 修, 棕田 亜砂美, 三好 真人
出版者	法政大学キャリアデザイン学会
雑誌名	生涯学習とキャリアデザイン
巻	18
号	1
ページ	11-29
発行年	2020-11
URL	<a href="http://doi.org/10.15002/00023625">http://doi.org/10.15002/00023625</a>

# チームワーク形成を目的とした PBL 型教育の効果測定： 研修評価アプローチによる検討

法政大学キャリアデザイン学部教授 安田 節之  
法政大学キャリアデザイン学部教授 梅崎 修  
サイボウズ株式会社 椋田亜砂美  
比治山大学現代文化学部講師 三好 真人

## 問題と目的

チームワークは、古今東西、馴染みの深い集団のスキルであり、現代社会では企業や組織といった集団における問題解決の一助になるとされている。仕事におけるタスク（task）が個人のキャパシティ（capacity）を超えてしまうケースはもとより、そもそもタスク自体が曖昧で、どの問題をどう解決してよいか分からないケース、仕事上のミス（例：“ヒヤリハット”）やエラー（例：ヒューマンエラー）を回避したいケース、さらには、複数メンバーで早急な意思決定が常に必要になるケースなど、チームワークが有効となる場面は多い（Salas, Cooke, & Rosen, 2008）。

個人がチームというプラットフォームで活動する機会が増えるなかで、チームワークによる問題解決の方法に注目が集まっている。実際、企業組織や教育機関で最も多く実施されるプログラムが Problem-Based Learning（PBL）型のチームワーク育成プログラムとされている（例：Beauchamp, McEwan, & Waldhauser, 2017; Klein, DiazGranados, Salas, Le, Burke, & Lyons, 2009）。他方、チームワーク育成プログラムのコンテンツや教育効果はまちまちであり、効果検証

が体系的に行われているものとそうでないものが混在している状況がある。そこで本研究では、高校生のチームワーク形成を目的として開発・実施された産学連携・PBL 型教育プログラム（サイボウズ社による「TPC プログラム」）に研修評価アプローチを用いた分析を行い、体系的な効果測定・評価研究のあり方を検討する。

## チーム研究

日常的に用いられるチーム（team）という概念は、学術的には、「タスクにおける相互依存が高く（強く）、共有化されかつ重んじられた共通のゴールをもつメンバーから構成された社会的存在」と定義される（Dyer, 1984 as cited in Salas, Cooke, & Rosen, 2008, p. 541）<sup>1</sup>。チームは集団での活動を想定しがちであるが、すべての活動がメンバー全員で行われるという訳ではない。メンバーが各自で取り組む活動であるタスクワーク（task work）、そして複数（または全員）のメンバーが協力して取り組む活動であるチームワーク（team work）の2つに大別されるのがチームである（例：相川・高本・杉森・古屋, 2012）。いずれの活動形式を取るにしても、メンバーが各々の立場からお互いを慮る行動をとり、ゴール達成

が目指されることになる。チームを定義するうえで「相互依存 (interdependence)」と「共通ゴール (common goal) の認識」が鍵となる所以はここにある (e.g., O'Brien, 1995)。逆に、各々が独立した存在として別々のゴールを目指すのであれば、そこにチームは成立し得ない。

チームビルディングとチームの有効性：これまで社会心理学や産業・組織心理学領域を中心に、様々な角度からチーム研究が行われてきた (例：山口, 2008；縄田・山口・波多野・青島, 2015)。一般に、チームに関する実践研究の蓄積は「どのようにチームが形成されるか (チームワークが醸成されるか)」というチームビルディング (team building)、そして「どのようにチームのパフォーマンスが向上するか」というチームの有効性 (team effectiveness) に大別できる。

まずチームビルディングは、ビルディング (building) という名が示す通り、介入 (intervention) を伴う概念で、企業などで最も採用されているアプローチである (e.g., Beauchamp et al., 2017; Klein et al., 2009)。具体的にチームビルディングは、広く一体感と団結を促進し、チームをよりスムーズに効果的に機能させるために設計されたものと定義される (e.g., Bruner, Eys, Beauchamp, & Côté, 2013)。つまり、単なる集団であるグループから相互依存と共通のゴールを有した“チーム”への変換を目指すのがチームビルディングである<sup>2</sup>。これまで、メンバー間の相互作用 (例：人間関係) の改善・向上を目的としたグループの構造や力学 (ダイナミクス) に立脚したチーム研究が行われてきた (Schein, 1969)。

一方、チームの有効性に関する研究では、どのようなチームの構造がパフォーマンス向上につながるかが理論化・モデル化されてきた。特に、Input-Process-Outcome (IPO) の枠組み (例：McGrath, 1964) では、①チームメンバーつまり個人レベルでの特性 (例：能力や性格)、チームレベルの要因 (例：リーダーシップやタスクの特

徴)、組織レベルの要因 (例：組織環境や組織風土) から構成される「インプット (input)」が、②ゴール達成に向けたメンバー間の相互作用、即ち「プロセス (process)」を生み出し、最終的に、③“質的・量的”ともに高いパフォーマンスや心理的な反応 (例：満足感やコミットメント) といった成果、つまり「アウトカム (outcome)」が導き出されるとされている。この IPO モデルを基本形とし、さらなるモデルの精緻化がなされており、例えば、Input-Mediator-Outcome (IMO) といったモデルに基づいてチームの有効性の検証が行われている (参考：Mathieu, Maynard, Rapp, & Gilson, 2008)。

チームエンパワメント：これまでの研究では、チームが目指すゴールやメンバーの役割分担が比較的明確なものが対象となることが多かった。スポーツチームであれば、監督やコーチのリーダーシップのもと、試合に勝利することが最終ゴールであるし、職場におけるチームであれば、そのチームを率いるリーダーを中心としたチームワークのもと業績 (例：利益) を上げることがゴールである。

一方、現代社会では、多様な社会的課題の解決を目的に、問題発見からゴール達成に至るまで、自らの力で行動するチームが求められている。昨今の学校や企業あるいは自治体では、そのような自律的なチームをどう形成しマネジメントしていくかに力点が置かれている。従って、チームのメンバーが自らを“エンパワー (empower)”して行動する、というチームエンパワメント (team empowerment) の発想が必要となってくるのである。

チームエンパワメントは、端的には文字通り、チームをエンパワーすることであるが、その根底には“パワー (power)”という複眼的な概念が含まれている。例えば、パワーは“対象に対する力 (power to)”であれば「影響力」となり、“対象を覆う力 (power over)”としては「支配力」となる。その一方で、“対象に関する力 (power of)”となれば「関係力」、さらに“ともに働く力

(power with) ”となれば「共働力」となる。これらはチームワークの基盤をなすパワーである。さらにパワーは人間関係における作用だけに留まらない。“うちにある力 (power within) ”となれば、個人における「内在力」ともなる (例：マクリーン・ハンソン, 2019)。

このように多面的なパワー概念を包含するエンパワメントは、あえて本質のみを捉えることを目的として、「個人・組織・コミュニティが自らの問題をコントロールするプロセスやメカニズム」と定義されている (Rappaport, 1987)。チーム研究におけるエンパワメント概念は、社会構造面 (social-structural) と心理面 (psychological) の2側面から考察が行われている (例：Seibert, Wang, Courtright, 2011; Mathieu, Gilson, & Ruddy, 2006)。

なかでも Mathieu et al. (2006) は、心理面に軸足を置いたチームエンパワメント (team psychological empowerment) を、「チームメンバーが、自らが近接する組織 (職場) 環境を自らがコントロールする権限を有し、チームの機能に対して自ら責任を持つという集団的な信念である」と定義している (p.98)<sup>3</sup>。また Kirkman and Rosen (1999) は、チームエンパワメントは次の4つの下位概念から構成されたとした：効能・潜在力 (potency)、意義・価値があるもの (meaningfulness)、自律性 (autonomy)、影響力 (impact)。さらに日本の文脈では、三島 (2007) によるエンパワメントの定義に即して<sup>4</sup>、チームエンパワメントが「チームの各メンバーが自らの力に気づいてそれを活用し、その力が個人・チーム内外の各層で展開するプロセスやメカニズム」と定義された (安田, 2018)。

以上の点に鑑みると、エンパワメント概念は現代社会が求めるチーム像、即ち、問題発見からゴール達成に至るまで自律的に活動できる集合体についての理解を深めるうえで、重要な役割を果たすと考えられる。つまり、チーム構造を理解する上では、既述した「相互依存」と「共通ゴールの認識」の他に、それらを下支えする基盤として「エンパ

ワメント」の視点を取り入れることが鍵となってくるのである。現代社会における質の高いチームを創造するうえで、エンパワメントが果たす役割は大きい (例：安田, 2014)。本研究では、エンパワメント概念を参考にしつつ、後述するチーム志向性 (team orientation) が産学連携プログラムによってどのように向上するかを検討する。

## 研修プログラムの効果測定とその問題点

教育機関における学びはもとより、企業組織や医療機関での人材育成や地域・自治体などでの生涯学習の場で、様々な研修プログラムが実施されている。これらの研修プログラムの中身は、問題解決 (problem solving)、教育・訓練 (education/training)、予防・促進 (prevention/promotion) に関することに大別することができる一方で (例：安田, 2011)、その効果測定については、実施されないケースや実施されても受講後の満足度アンケートなど、方法論上、不十分なケースがほとんどである。

例えば、参加者の事後の感想・コメントや受講の満足度をもとに研修プログラムの効果測定を行った結果、研修プログラムで有意義な時間を過ごし、満足できた参加者がいたとする (例：「とても満足である」と研修を評価)。しかし、研修に満足できた一方で、そもそもの意義ある学び (例：知識やスキルの習得) が起こっていなかった場合、その研修は成功したとは必ずしも言えない<sup>5</sup>。

研修の企画者や実施主体などのステークホルダーは、つまるところ、受講者 (あるいは組織) に何らかの変化 (“どの程度の効果”) が認められたのかについての情報を欲している。しかし、このシンプルな問いを科学的に探究するためには、研修と効果の間に存在する因果関係を設定・実証する必要がある、そのためには、事前・事後における測定はもとより比較群の設置や無作為化の実施などが必要となってくるのである<sup>6</sup>。

また効果と一口に言っても、何をどう推し量るのか、という概念設定や測定上の課題もある。

研修プログラムを実施するにあたっては、組織などが抱える問題解決への即時性や費用対効果が求められていることが多い。その場合、厳格な (rigorous) 科学的手法に則った効果測定が実施できるような時間的・経済的余裕を有していることはほとんどない。つまり、科学的に理想とされる研修評価アプローチを用いた効果測定の適用は、残念ながら、実際には極めて困難となるのである。このような研修評価に関する問題点をもとに、簡易かつ体系的な効果測定として開発されたのがカークパトリック (D. Kirkpatrick) による4段階アプローチ (FLA: 4-Level Approach; Kirkpatrick & Kirkpatrick, 2006) である。

### カークパトリックの4段階アプローチ (FLA: 4-Level Approach)

研修プログラムはどのようにすれば体系的・実証的に評価することができるのか。この問いを突き詰めて開発された研修評価アプローチがカークパトリックによるFLAである (Kirkpatrick & Kirkpatrick, 2006)。

FLAは、評価には様々なステークホルダーが存在し評価を行う際には様々な立場や意見を取り入れる必要がある、という基本姿勢をとっている。例えば、研修後における受講生の行動変容によってその研修の良し悪しが判断 (評価) されるべきという立場がある。一方、研修後の職務上の成果のみを研修評価の対象とすべきという見方もある。さらには、研修後に収集される感想・コメントや満足度の評定が研修評価において欠かせないとする立場もある。または研修を受けるなかで受講生が習得する知識・態度・スキルこそが研修を受けることの意義・価値であるとする立場もある (Kirkpatrick & Kirkpatrick, 2006)。

Kirkpatrick & Kirkpatrick (2006) はこれらのすべての立場が重要であるとしたうえで、評価の対象から方法に至るまでの体系的な情報集約をFLAとして行うことの可能性を提示した。即ち、研修を受ける側 (例: 受講者・学習者) の①「反応 (reaction)」、②「学習 (learning)」、③「行

動 (behavior)」、④「結果 (results)」という4段階 (4-level) における変化についての測定・評価である。従って、Kirkpatrick & Kirkpatrick (2006) のFLAでは、まず研修プログラムの参加者 (学習者) への影響が次の4段階から捉えられることになる (参考: 安田, 2018) :

- Level 1: 「反応 (reaction) レベル」学習者は研修に満足したか。
- Level 2: 「学習 (learning) レベル」学習者に学びはあったか・知識等を習得したか。
- Level 3: 「行動 (behavior) レベル」研修後、学習者の行動が変化したか。
- Level 4: 「結果 (results) レベル」学習者を取り巻く環境が肯定的に変化したか。

反応レベル: 反応 (reaction) レベルはFLAにおける第一段階に位置づけられる。この段階では、教育・研修プログラムを受けての受講者の感想や満足度について、それらが肯定的なものであったのか、あるいは否定的なものであったのかが評価される。反応レベルの評価は、教育・研修プログラムにおける一般的な評価形式として数多く行われている。もし反応レベルが否定的な結果を伴っていた場合には (例: 研修に満足できなかった)、当然、次にみる学習レベルの高さは期待できない。しかし、仮に反応レベルに肯定的な結果であったとしても (例: 研修に満足できた)、それだけで学習レベルの高さが保証される訳ではない (例: 知識やスキルが習得できた) (Kirkpatrick & Kirkpatrick, 2006)。

反応レベルにおける評価基準や評価項目は研修によって異なるが、具体的な評価基準としては「研修に満足したか」「役に立つ研修であったか」「楽しく受講することができたか」「研修内容を理解できたか」「自分が興味を持っていた内容であったか」「研修内容についてもっと知りたいと思ったか」「新たな気づきが得られたか」といった例が考えられる (Yasuda et al., 2018; 安田, 2018)。



学習レベル：FLAにおける第2段階は「学習 (learning) レベル」の評価である。ここでは、学習者が研修で何をどの程度学んだか、即ち、知識やスキルの習得の有無や多寡が評価の対象となる。プログラムの実施前後において習得した知識の量とともに、学習者の態度やスキル、興味・関心の変化 (cf. 非認知能力) などについての評価も行われることもある。これらは知識 (Knowledge)、態度 (Attitude)、スキル (Skill)、興味・関心 (Interest) の頭文字をとって「カシー (KASI)」と呼ばれ心理・教育測定における効果の類型化として用いられる要因に他ならない (参考：安田, 2013)。

一般に、教育・研修プログラムでは予め達成目標・到達目標などが掲げられるが、それらは、ここで言うカシーの枠組みから捉えられるものが殆どである。したがって、学習者の知識・態度・スキル・興味などの変化をプログラムにおける到達目標および行動レベルの評価対象と関連づけながら、学習レベルの評価を行っていくことが求められる。

行動レベル：教育・研修プログラムを受講した後に学習者に起こる行動の変化が「行動 (behavior) レベル」の評価である。プログラムのステークホルダーである人材育成などを依頼する側やプログラムの運営側が知りたいのは、研修後に社員・従業員の行動に望ましい変化があったか否かである。そのため、事後的に生じた行動の変化のみに着目し、評価が行われることも多い。

しかし、これまでみてきた「反応レベル」と「学習レベル」の評価を行わずに、一足飛びに「行動レベル」の評価を行うことは避けられるべきであるとされている (Kirkpatrick & Kirkpatrick, 2006)。少なくとも下記の4つの段階を経てはじめて、行動レベルに変化が起こるからである (Kirkpatrick & Kirkpatrick, 2006, p. 23)：

- ①学習者が実際に (行動レベルの) 変化を望んでいる。
- ②何を (what)、どのように (how) すれば変

化できるかを学習者が分かっている。

- ③学習者が適切な環境 (right climate) にいる。
- ④行動を変化すれば、学習者にやりがい (rewarding) や報酬がもたらされる。

これらの段階は、教育・研修プログラムのなかだけに留まらず、事後的な組織・環境要因 (例：学習や労働環境) に左右されるものである。したがって、いかに教育・研修プログラムの内容や目的が、受講者・学習者はもとより、その他の多くのステークホルダーに共有されているか、事後的にどのように行動変容に向けたサポートが行われるか、がプログラムを成功に導くための鍵となる。

変化・変容を促したいと考える行動は、時にプログラムのアウトカム (outcome) として捉えられる。そして、そのアウトカムの数は、研修の数だけ存在することになる。ここですべてのアウトカムは検討できないが、例えば、職場におけるチームワーク向上プログラムであれば「チームの目指す方向を意識して行動しているか」「自分の長所がチームメンバーから認められているか」「チームにおいて自分と異なる意見にも耳を傾けるか」といった行動レベルのアウトカムが想定できる (Yasuda, Umezaki, Mukuta, & Miyoshi, 2018；安田, 2018)。行動レベルの評価には、次にみていく結果レベルの評価と同様に、因果関係の特定・実証が必要となる。

結果レベル：FLAにおける最後の段階が「結果 (results) レベル」の評価である。研修プログラムという「原因 (cause)」によってどのような「成果 (effect)」が生まれたのか、という因果関係がこの結果レベルの評価で明らかにされる。教育・研修後に学習者の成績が上がったのか、各種の領域において何らかの成果があったか (例：受賞実績)、ビジネスにおける研修であれば、従業員の業績や効率の向上・改善につながったか、多様なコスト (例：時間的・経済的) を削減することができたのかなどがこの段階での評価の対象となる。

結果レベルの評価は、プログラムの存在価値の

試金石となる一方で、評価にあたってはいくつかの困難が伴う。まず、結果を正確に定義・測定すること、つまり何をもって成績や業績の向上や仕事効率の改善とするかなどを、教育・研修プログラムのコンテンツや方法と照らし合わせ、体系的に把握することは難しい。またプログラム以外の別の影響が常に存在するため、得られた結果が真に教育・研修プログラムによるものなのかが不透明であることが多い。しかしエビデンス（科学的根拠）を正確に提示することは即ち、原因（教育・研修プログラム）と結果（受講生の成果・効果）との間に確かに因果関係が成立することを実証するに他ならない。他のレベル（例：反応・学習）にも増して、評価デザインに注意を払う必要がある理由はここにある（例：安田，2018）。

体系的な研修評価アプローチの必要性：以上、カークパトリックのFLAについて検討してきた。一般に、教育・研修プログラムの企画・実施主体（ステークホルダー）の主要な関心事は、学習者への成果・効果（アウトカム）の有無や多寡である。これは即ち、行動レベル（Level 3）の変化に対する問いに相当する（ie., 研修後に行動が変わったか）<sup>7</sup>。しかし、この問いの科学的な検証には実験デザインの整備を要するため（例：比較群の設置や無作為化の実施）、現実的ではないことが多い（参考：安田，2018；安田，2011）。

しかし一方で、FLA（Kirkpatrick & Kirkpatrick, 2006）のように「反応レベル（Level 1）」と「学習レベル（Level 2）」の評価を「行動レベル（Level 3）」の評価と併せて行い、実験デザインとは別の角度から、体系的な研修評価を行っていくことも可能である。本研究ではそのような研修評価アプローチの可能性を追究する。

## 本研究の目的

様々な領域でチームによる問題解決の重要性が指摘され、チームワーク育成プログラムの開発や実施が増えるなか、効果測定・評価研究のあり方が問われている。プログラムの効果測定では、学

習者が複数回にわたるプログラムのなかで、何を感じ取り、どのように肯定的な行動につながったのかを、システマティック（体系的）に評価する必要がある。そこで本研究では、カークパトリックのFLA（Kirkpatrick & Kirkpatrick, 2006）のうちの3段階、即ち、反応レベル（level 1）、学習レベル（level 2）、行動レベル（level 3）における研修評価アプローチを用いた効果測定を行う。具体的には、問題解決のためのチームワーク育成を目的とした高校生を対象とした産学連携・PBL型の教育プログラム（詳しくは、椋田・梅崎・安田・三好，2018を参照）の効果測定にFLAを応用し、より体系的な研修評価アプローチのあり方を検討することを目的とする。

## 研究 1

### 目的

研究1の目的は、産学連携・PBL型教育プログラムであるTPCプログラムの実施状況や介入課程のモニタリングを行うことであった。具体的には、カークパトリックのFLAに基づく反応レベル（Level 1：学習者（生徒）は意欲的に学んでいたか）および学習レベル（Level 2：生徒は知識・スキルを習得していたか）においてプロセス指標を設置し効果測定を行った。

### 方法

TPC プログラム：TPC プログラムは、チームによる問題・課題解決の方法を学ぶことを目的とした計5回の講義（セッション）から成っている。“TPC”即ち、チームワーク（Teamwork）、問題解決（Problem-solving）、コミュニケーション（Communication）を学びの柱とし、計5回（各回2時間）のセッションでは、①ゲスト講師による講演、②チームワークや問題解決メソッドについて、③企画書の作り方、まとめ方、プレゼンテーションの方法、④チームの進捗チェック、⑤最終発表会から構成されていた（詳細のプログラムの実施報告は、椋田ほか、2018を参照）。

調査対象と手続き：2017年5月から11月にかけて、神奈川県の高校（介入群 A）および北海道の高校（介入群 B）の合計2つの高校で TPC プログラムが実施された。介入群 A（N=126）は女子校であった。他方、介入群 B（N=90）は共学校であった。また介入群 A と同一高校において、TPC を受講しないクラスである比較群（N=118）からのデータ収集も併せて行った。比較群には TPC プログラムの代わりに、異文化理解に関する授業が実施された。

TPC プログラムは計5回から構成されており、介入群 A および介入群 B については毎回のプログラム実施後、反応レベルおよび学習レベルの測定が実施された（i.e., 時期1～時期5）。また比較群については、プログラム実施開始（i.e., 時期1）および終了時（i.e., 時期5）の2時点においてのみ反応レベルの測定が行われた。

プロセス指標の開発と有効化：反応レベルおよび学習レベルの測定を行うために、それぞれのプロセス指標が開発された。まず反応レベルでは「生徒は意欲的に学んでいたか」という評価（リサーチ）クエスションのもと計12項目から成るプロセス指標が作成された（Appendix A）。具体的には、受講内容への興味や満足度、グループワークでのコミュニケーションに関する項目から構成された。これらの反応レベルに関する項目に対して、「そう思う（5）」「ややそう思う（4）」「どちらとも言えない（3）」「あまりそう思わない（2）」「そう思わない（1）」の5件法による回答が求められた。

反応レベルの信頼性係数（Cronbach's alphas）は、介入群 A では .90（時期1）、.93（時期2）、.93（時期3）、.94（時期4）、.95（時期5）、介入群 B では .79（時期1）、.93（時期2）、.93（時期3）、.93（時期4）、.91（時期5）となっていた。比較群における反応レベルの測定は初回（時期1）および最終回（時期5）のみ実施され、それぞれ、.85および.81となっていた。

次に学習レベルでは「生徒は知識・スキルを習

得していたか」という観点からプロセス指標が作成された<sup>8</sup>。プログラムで学ぶ内容は毎回異なるため、受講内容の理解の度合いについての簡単な客観テスト（小テスト）が各回のプログラム実施後に行われた。小テストは、各回の重要ポイントについて問われたものであり、作問および採点はプログラム実施者（講義担当者）によって行われた。問題数・得点範囲も各回において異なっていたため授業理解度としてパーセンテージ（%）表示がなされた（表1の学習レベルを参照）。

## 結果と考察

反応レベル：反応レベルの平均値（SD）を表1に示した。介入群の種別および時期の変化による反応・学習の違いを検討するために、2元配置反復測定分散分析（Two-way Repeated Measures ANOVA）を行った。その結果、交互作用  $F(1, 135)=10.97, p < .01, \eta^2 = .07$ 、時期要因（5レベル）の主効果  $F(1, 135)=25.50, p < .01, \eta^2 = .15$  および群要因（A および B）の主効果  $F(1, 135)=24.37, p < .01, \eta^2 = .15$  が認められた。

反応レベルは時期1が他の時期に比べて低かった半面、それ以降、時期2、時期4、時期5と高い水準で推移していた（時期の主効果）。反面、時期3の反応レベルについては、他の時期に比べて低かった。また介入群要因については、いずれの時期においても介入群 A が介入群 B よりも高い平均値で推移していた（群の主効果）。データ全体では、交互作用が認められたものの、介入群 A・B ともに各平均値は4.00（「ややそう思う」）に近い値を示していたため、一部の例外を除き（例：介入群 B の時期1）、どちらの介入群も TPC プログラムに満足し意欲的に取り組んでいたと判断できる。

次に TPC プログラム（介入群 A）と非 TPC プログラムである他の授業（比較群）の反応レベルを比較するために、同一高校である介入群 A と比較群における時期1と時期5の2地点データを分析した<sup>9</sup>。その結果、両者の反応レベルには有意な交互作用が認められた  $F(1, 210)=26.45,$



表1 2元配置反復測定分散分析 (Two-way Repeated Measures ANOVA) の結果

	時期 1		時期 2		時期 3		時期 4		時期 5		群	時期	交互作用
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	F 値	F 値	F 値
反応レベル											24.37*	25.50*	10.97*
介入群 A	3.95	0.64	4.26	0.63	3.89	0.65	4.35	0.58	4.22	0.58			
介入群 B	3.20	0.60	3.95	0.72	3.85	0.63	3.73	0.74	3.80	0.63			
学習レベル											14.24*	15.97*	11.14*
介入群 A	89.75	12.75	86.54	18.81	68.15	18.64	95.56	16.15	87.78	24.71			
介入群 B	81.23	13.74	90.63	11.19	71.32	33.10	72.55	37.83	68.63	38.54			
行動レベル											31.18**	9.23*	.44
介入群 A	5.58	0.92	5.85	0.73	5.73	0.87	5.93	0.82	6.04	0.73			
介入群 B	4.90	0.96	4.98	0.98	4.92	1.11	5.23	1.17	5.32	1.02			

注1：要因1（介入群 A・B）、要因2（時期1～時期5）。Mean: 推定周辺平均（Estimated Marginal Mean）。\* $p < .01$

注2：反応レベルのサンプル数は  $N = 90$ （介入群 A）および  $N = 47$ （介入群 B）、学習レベルのサンプル数は  $N = 90$ （介入群 A）および  $N = 51$ （介入群 B）、行動レベルのサンプル数は  $N = 77$ （介入群 A）および  $N = 45$ （介入群 B）であった。

$p < .01$ ,  $\eta^2 = .11$ 。その一方で、時期要因 [ $F(1, 210) = 52$ ,  $p = \text{n.s.}$ ,  $\eta^2 = .00$ ] および介入・比較群要因 [ $F(1, 210) = 1.70$ ,  $p = \text{n.s.}$ ,  $\eta^2 = .01$ ] における単純主効果は認められなかった（図1）。これは介入群・比較群の反応レベルが時期によって受ける影響が異なっていたことを示している。

時期1では比較群における反応レベルが介入群 A よりも高かったが、時期5ではその傾向が逆転し、今度は、介入群 A の反応レベルが比較群よりも高かった。したがって、TPC プログラム（介入群 A）が開始時の学びの意欲や授業満足度はその他の授業（比較群）よりも低いものの、終了時では逆にプログラムでの学習意欲・授業満足度がその他の授業よりも高い状態となっていたことが明らかになった。

学習レベル：介入群の種別および時期の変化による学習レベルの違いを検討するために、2元配置反復測定分散分析（Two-way Repeated

Measures ANOVA）を行ったところ、交互作用 [ $F(1, 139) = 11.14$ ,  $p < .01$ ,  $\eta^2 = .07$ ]、時期要因（5レベル）の主効果 [ $F(1, 139) = 15.97$ ,  $p < .01$ ,  $\eta^2 = .10$ ] および介入群要因（A および B）の主効果 [ $F(1, 139) = 14.24$ ,  $p < .01$ ,  $\eta^2 = .09$ ] が認められた（表1）。群および時期による学習レベルへの影響としては、介入群 A における時期3の理解度が他の時期よりも比較的低く（ $ps < .05$ ）、かつ時期4・時期5において特に介入群 A が介入群 B よりも理解度が高かった（ $ps < .05$ ）<sup>10</sup>ということ以外（表1参照）、解釈可能なパターンは存在しなかった。各時期・群ともに、70%から90%程度の得点率となっていたことより、プログラムでの学習レベルは良好であったと考えられる。

一方、表2の相関係数表をもとに、学習レベルが反応レベルとどのように関連していたかを検討したところ、一概に高い相関関係があったとは言えないことも明らかになった。両者の相関が高い、

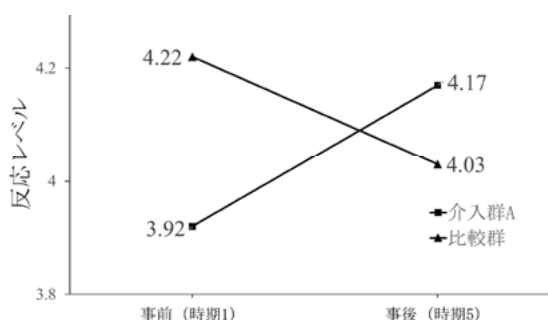


図1 反応レベルの変化：不等価（介入群・比較群）事前事後テストデザインによる検討

即ち、授業への意欲・満足度（反応レベル）が高ければ高いほど、理解度（学習レベル）が高かったのは、介入群 A の時期 1 ( $r = .74, p < .01$ ), 時期 3 ( $r = .48, p < .01$ ), 時期 5 ( $r = .46, p < .01$ ) のみであった。逆に、介入群 A の時期 2 ( $r = .04, p = n.s.$ ), 介入群 B の時期 2 ( $r = -.17, p = n.s.$ ), 時期 4 ( $r = .12, p = n.s.$ ) および時期 5 ( $r = .09, p = n.s.$ ) については、学習レベルと反応レベルとの間の相関は低かった。したがって、これらの時期については、講義に楽しく、意欲的に取り組めた一方で、必ずしも知識やスキルの習得にはつながっていなかったとも考えることができる。

## 研究 2

### 目的

研究 2 の目的は、TPC プログラムのアウトカム評価を行うことであった。具体的には、カークパトリックの 4 段階アプローチに基づくレベル 3 における「研修後、学習者の行動が変化したか」(行動レベル) というリサーチクエスションに基づく効果測定が行われた。

### 方法

調査対象と手続き：研究 1 における調査対象と手続きのとおり同時期・同様形式にてデータ収集が行われた。

アウトカム指標（行動レベル）の開発と有効化：「行動レベル」でのアウトカム指標として、チーム志向性尺度が用いられた。この尺度は、これまでみてきた高校生のチーム開発を目的として設計・実施された産学連携・PBL 型教育である TPC プログラムの効果測定研究の一環として開発されたものである（例：Yasuda et al., 2018; 安田, 2018）。研究 1 のプロセス指標（反応レベル・学習レベル）による測定に続き、研究 2 ではアウトカム指標（行動レベル）による測定が行われたが、ここで使用されたのがチーム志向性（team orientation）尺度である<sup>11</sup>。

チーム志向性尺度は合計 10 項目から構成されており、チームワークの基本とされるメンバー間の相互依存と共通ゴールの認識（e.g., O'Brien, 1995）が主な設問の内容であった。具体的には、「自分の長所がチームメンバーから認められている」「自分で分からないことがあれば、他のチームメンバーに聞くことができる」といった項目によってメンバー間の相互依存が、「チームの目指す方向を意識して行動する」「チームが何らかの成果・結果を出せるように、協力するほうだ」といった項目によって共通ゴールの認識が測定された。これらの項目に対して「とてもそう思う (7)」から「まったくそう思わない (1)」までの 7 件法による回答が求められた。

表 2 時期 1 から時期 5 における反応・学習・行動レベルの相関係数

	時期 1			時期 2			時期 3			時期 4			時期 5		
	反応	学習	行動	反応	学習	行動	反応	学習	行動	反応	学習	行動	反応	学習	行動
時期 1															
反応レベル	1	.33**	.32**	.31**	.10	.14	.32**	.20*	.29**	.28**	.12	.30**	.33**	.11	.35**
学習レベル	.74**	1	.05	.35**	-.04	.22*	.15	-.02	.09	.11	-.02	.01	.14	-.01	.13
行動レベル	.37**	.56**	1	.32**	-.06	.31**	.32**	.13	.45**	.19	-.10	.21*	.16	-.01	-.07
時期 2															
反応レベル	.49**	.48**	.50**	1	-.17	.51**	.34**	.01	.30**	.36**	-.09	.27**	.27**	.07	.26**
学習レベル	-.10	0.03	-.10	.04	1	.09	.00	.18	.06	.03	.30**	.13	-.03	.20*	-.02
行動レベル	.35**	.45**	.72**	.73**	.11	1	.25**	-.02	.29**	.25*	.06	.36**	.30**	.15	.27**
時期 3															
反応レベル	.42**	.51**	.29*	.66**	-.02	.54**	1	.28**	.59**	.50**	.10	.39**	.51**	.18	.28**
学習レベル	.14	.16	.08	.29*	.15	.23	.48**	1	.34**	.18	.17	.31**	.27**	-.10	.32**
行動レベル	.30**	.45**	.64**	.66**	.11	.82**	.60**	.48**	1	.45**	.02	.57**	.50**	.00	.46**
時期 4															
反応レベル	.47**	.34**	.38**	.69**	-.08	.56**	.55**	.32*	.57**	1	.12	.69**	.67**	.14	.56**
学習レベル	.33*	.23	.22	.58**	-.01	.40**	.51**	.51**	.42**	.39**	1	.00	.16	.33**	.12
行動レベル	.49**	.39**	.53**	.77**	-.08	.75**	.53**	.27*	.72**	.85**	.37**	1	.62**	.06	.65**
時期 5															
反応レベル	.38**	.49**	.35**	.53**	.01	.54**	.68**	.38**	.52**	.60**	.26*	.64**	1	.09	.68**
学習レベル	.40**	.38**	.35**	.44**	.09	.50**	.35**	.11	.40**	.42**	.40**	.49**	.46**	1	.03
行動レベル	.47**	.53**	.65**	.69**	-.01	.77**	.55**	.27*	.76**	.72**	.29*	.81**	.65**	.51**	1

注 1：行動レベル（チーム志向性）

注 2：上段（介入群 A）、下段（介入群 B）

因子分析と信頼性係数：チーム志向性尺度の因子構造を検討するために、複数の介入群・統制群および介入時期のデータへの探索的因子分析（EFA：Exploratory Factor Analysis）を行った。最尤法（maximum likelihood）による推定で固有値1以上を因子抽出の基準とし、スクリープロットによる固有値の増減を考慮し判断したところ、すべてのデータにおいて一因子の抽出が妥当であることが明らかになった（i.e., 介入群 A・B は時期1～時期4のデータ、比較群は時期5データ）。

表3に示された通り、因子負荷は.52から.93までとなっていた一方で、多くが.70sから.80sと十分な値を示していた。また当該因子によって全体の48.93%から72.10%の分散が説明されていた。以上より、本尺度は1因子構造となっていたことが確認された。

次に、EFAによる1因子構造が確証的因子分析（CFA：Confirmatory Factor Analysis）によっても実証できるかを確かめるために、プログラム実施後の別データ（介入群の時期5データ）への多母集団確証的因子分析（MGCF A：Multi-Group Confirmatory Factor Analysis）を行った。その結果、EFAと同様に、1因子構造によるデータ適合が確認された（例：CFI=.91, NFI=.89, RMSEA=.09）。表3に示したとおり因子負荷も高い値を示していた。

最後に、尺度の信頼性を確かめるために、ク

ロンバックの $\alpha$ 係数（Cronbach's alpha）を算出したところ、.88から.95と十分な値をとっていた（Yasuda et al., 2018）。以上より、チーム志向性尺度は一定の構成概念妥当性（因子構造）と信頼性を有していたことが確認された。

効果測定の実デザイン：TPCプログラムの効果測定（アウトカム評価）は、2つのデザインによって行われた。最初に、プログラム（介入群 A・介入群 B）による効果、即ち学習者の行動変容（i.e., チーム志向性の向上）がどのようなプロセスで起こっていたか、についての検討を行った。そして次のフェーズでは、当該プログラム（介入群 A）は他のプログラム（比較群）と比べてどの程度、効果があったか、についての評価が行われた。

前者の分析には図2に示したモデルについての多母集団潜在成長モデリング（MGLGM：Multi-Group Latent Growth Modeling）、後者の分析には不等価統制群事前事後テストデザイン（non-equivalent control design）が用いられた。最後に、反応および学習レベルの分析と同様に、行動レベルにおいても介入群の種別と時期による違い（変化）が認められるかの分析が行われた（i.e., 反復測定分散分析）。

## 結果と考察

まず MGLGM による分析について、図2の基本モデルにおける切片（level）および傾き（slope）

表3 チーム志向性尺度の探索的・確証的因子分析の結果

チーム志向性	因子負荷 [EFA]			因子負荷 [CFA]		
	介入群A	介入群B	比較群	介入群A	介入群B	比較群
1. チームの目指す方向を意識して行動する。	.76	.86	.84	.80	.89	.82
2. チームが何らかの成果・結果を出せるように、協力するほうだ。	.79	.86	.84	.88	.88	.81
3. 自分の長所がチームメンバーから認められている。	.52	.78	.68	.78	.77	.68
4. チームでの自分の役割は明確だ。	.67	.81	.65	.81	.79	.71
5. チームでは自分の役割を、怠けることなく果たしている。	.67	.87	.69	.86	.88	.72
6. チームでの話し合いのときに、自分の意見をしっかりと言うほうだ。	.62	.88	.61	.83	.77	.66
7. 自分で分からないことがあれば、他のチームメンバーに聞くことができる。	.71	.82	.75	.80	.89	.59
8. 自分が勝手に行動をとって良いか分からないときには、事前に他のチームメンバーに相談する。	.75	.91	.79	.81	.85	.55
9. チームでの話し合いのなかで意見の食い違いがあった場合、自分と異なる意見にも耳を傾けるほうだ。	.76	.85	.80	.86	.83	.59
10. 結果を出すためにチームで「やること」を決められる。	.79	.83	.78	.85	.71	.68

注）EFA: Exploratory Factor Analysis [介入群 A・B：時期1～時期4のデータ；比較群：時期5]；CFA: Confirmatory Factor Analysis [時期5データ]

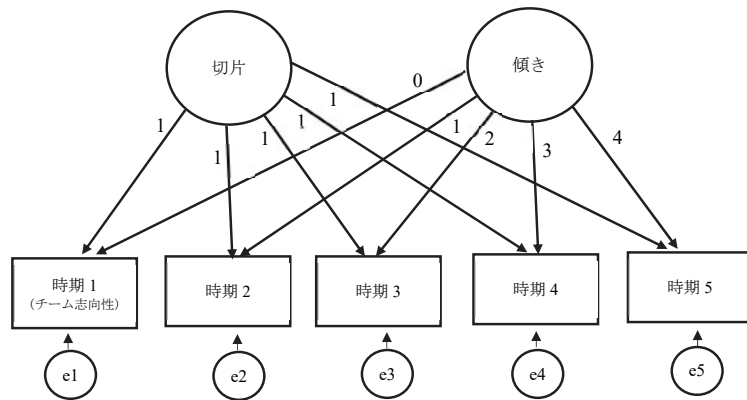


図 2 チーム志向性の潜在成長モデリング

の相関をコントロールしたモデルの分析を行った結果、モデル適合度は十分であった [ $\chi^2(20)=46.50$ ,  $p<.01$ ; the comparative fit index (CFI) = .93; the normed fit index (NFI) = .89; and the root mean squared error of approximation (RMSEA) = .07]。モデル中のチーム志向性の切片 (level = 平均値) は介入群 A・B ともに高い水準で推移しており、介入群 A > 介入群 B となっていた。また両群ともに切片の分散に有意な結果が認められた (表 4)。

また、チーム志向性の変化を示す傾き (slope) に関しても、介入群 A・B ともに有意な結果が認められた (表 4)。介入群 A では傾きの分散も有意な値を示していた一方で、介入群 B にはそのような結果は認められなかった。したがって、介入群 A よりも介入群 B のほうが、比較安定した肯定的変化があったことが明らかになった。また、切片と傾きの相関係数は介入群 A は  $r = -.62$  ( $p<.01$ )、介入群 B は  $r = .05$  ( $p=n.s.$ ) となっていた。

以上より、介入群 A については、チーム志向性が開始時に低かった受講者にとってはプログラムを受けることで確かにチーム志向性を高められていた反面、もともとチーム志向性が高かった受講者は、プログラムを受講することにより逆にそれを低めてしまっていた。一方、介入群 B については、もともとのチーム志向性の高低はプログラムによる変化には影響を及ぼさず、回を追うごとに、着実にチーム志向性が高まることが明らかになった。

次に、行動レベルは、FLA における前段階である反応レベルや学習レベルによる影響を受けるか、という点について検証するために、先にみた基本モデル (図 2) に、時期 1 から時期 5 までの反応レベルと学習レベルを外生要因として新たに組み込んだモデルを作成し、MGLGM による分析を行った。

その結果、反応レベルおよび学習レベルともに確かに何らかの影響を及ぼすことが明らかになっ

表 4 潜在成長モデリング (LGM) における切片 (Level) と傾き (Slope)

	介入群 A		介入群 B	
	推定値	分散	推定値	分散
切片	5.69** (.07)	.37** (.08)	4.98** (.10)	.76** (.14)
傾き	.06* (.027)	.06** (.01)	.10** (.02)	.01 (.00)

注：切片 (Level)、傾き (Slope)；介入群 A (N=)、介入群 B (N=)。\*\* $p<.01$ , \* $p<.05$



た（表5）。具体的には、介入群 A については、各回の行動レベルの「高低（切片≒平均値）」が、特に時期2（ $\beta = .61$ ,  $p < .01$ ）と時期3（ $\beta = .52$ ,  $p < .01$ ）の反応レベルの高低に強く影響されていたことが明らかになった。一方、行動レベルの「変化（傾き）」については、時期4（ $\beta = .44$ ,  $p < .01$ ）と時期3（ $\beta = .46$ ,  $p < .01$ ）における反応レベルの影響が大きかった。しかし学習レベルについては、行動レベルの高低や変化に特に影響を与えていなかったことも明らかになった。

以上より、介入群 A については、時期2と時期3におけるプログラム満足度が高かった受講者は全体的にそもそもの行動レベルが高い人たちであった一方で、時期4と時期5におけるプログラム満足度が高かった受講者は、プログラムを受講することにより行動レベルをより一層高められていた人たちであったと言える。また、切片と傾きの逆相関がすべての時期において強かったことより（rs: -.67から-.83;  $ps < .01$ ）、先にもみたとおり、当該プログラムは行動の基本レベルがより低い受講者により肯定的な変化をもたらす一方で、行動レベルの高い受講者にはむしろ逆効果であることも明らかになった。

次に、介入群 B であるが、ここでは反応レベルと学習レベルの影響はより強く表れていた（表5）。つまり、行動レベルの「高低（切片）」は、時期2（ $\beta = .73$ ,  $p < .01$ ）、時期3（ $\beta = .51$ ,  $p < .01$ ）、時期4（ $\beta = .58$ ,  $p < .01$ ）の反応レベル、そして行動レベルの「変化（傾き）」は、時期3（ $\beta = .47$ ,  $p < .01$ ）、時期4（ $\beta = .89$ ,  $p < .01$ ）、時期5（ $\beta = .58$ ,  $p < .01$ ）の反応レベルにそれぞれ影響を受けていた。また学習レベルについても、時期1（ $\beta = .58$ ,  $p < .01$ ）における学習レベルが高かった受講者は全般的な行動レベルが高かった人であった一方で、時期2（ $\beta = .46$ ,  $p < .01$ ）における学習レベルが高かった人ほど、プログラム全般において行動レベルを逆に低めてしまっていたことも明らかになった。さらに、時期2（ $r = -.33$ ,  $p < .01$ ）および時期5（ $r = -.34$ ,  $p < .01$ ）において行動レベルの切片と傾きには逆相関が認められたものの、全時期においてより強い逆相関を示した介入群 A と比べて、その関連性は限定的であった。

続いて、不等価統制群事前事後テストデザインに基づきプログラム実施群（介入群 A）と非実施群（比較群）から収集されたデータについて反復測定分散分析（Repeated Measures ANOVA）

表5 反応レベルおよび学習レベルが行動レベルの切片（level）・傾き（slope）に及ぼす影響

		反応レベル		学習レベル		切片・傾き	モデル適合度			
		切片	傾き	切片	傾き		$\chi^2/df(20)$	CFI	NFI	RMSEA
時期1	介入群 A	.26*	.10	.01	.00	-.70**	2.57	.90	.86	.08
	介入群 B	-.04	.61**	.58**	-.17	-.10				
時期2	介入群 A	.61**	-.15	.21	-.08	-.67**	2.45	.90	.86	.08
	介入群 B	.73**	.38*	.10	-.46*	-.33*				
時期3	介入群 A	.52**	-.11	-.02	.25*	-.69**	2.59	.89	.85	.08
	介入群 B	.51**	.47*	.00	.12	-.23				
時期4	介入群 A	.28*	.44**	-.16	.10	-.79*	2.49	.91	.86	.08
	介入群 B	.58**	.89**	.06	.01	-1.11**				
時期5	介入群 A	.22	.46**	.02	-.01	-.83**	1.60	.96	.91	.05
	介入群 B	.38**	.58**	.35**	-.03	-.34*				

注1) 多母集団潜在成長モデリング（MGLGM：Multi-Group Latent Growth Modeling）による推定。

注2) 切片・傾き（共に行動レベル）への影響、反応レベルおよび学習レベル（回帰係数 $\beta$ ）、切片・傾き（相関係数 $r$ ）

注3) 時期4（介入群 B）における推定値（ $r = -1.11$ ,  $p < .01$ ）は多重共線性等の影響も考えられたため解釈を行わなかった。

注4) \* $p < .05$ , \*\* $p < .01$

を行った結果（図3）、交互作用が認められた  $[F(1, 196) = 5.12, p < .01, \eta^2 = .02]$ 。また時期  $[F(1, 196) = 8.78, p < .01, \eta^2 = .04]$  および群  $[F(1, 196) = 9.41, p < .01, \eta^2 = .04]$  の単純主効果も認められた。

Bonferroni による多重比較検定の結果、チーム志向性が統計的に有意に事後（時期5）＞事前（時期1）となっていたのは介入群 A のみであったと同時に、事前（時期1）の段階では介入群 A・比較群に有意差は存在しなかったが、事後（時期5）の段階ではチーム志向性が介入群が比較群よりも有意に高くなっていた ( $ps < .01$ )。以上より、プログラムの非実施群と比べた場合、TPC プログラムの実施がチーム志向性の向上に有効であることが確認された。

最後に2元配置反復測定分散分析（Two-way Repeated Measures ANOVA）を行い、介入群の種別や時期によって行動レベルに違いが生じていたかを検討した結果、交互作用  $[F(1, 120) = .44, p = n.s., \eta^2 = .00]$  は統計的に有意ではなかったが、時期要因（5レベル）の主効果  $[F(1, 120) = 9.23, p < .05, \eta^2 = .07]$  および介入群要因（A および B）の主効果  $[F(1, 120) = 31.18, p < .01, \eta^2 = .20]$  に有意な結果が認められた（表1）。

群および時期による行動レベルへの影響として特徴的な点としては、時期5がどの時期と比べても行動レベルが高かった点 ( $ps < .05$ )、いずれの時期の行動レベルも介入群 A > 介入群 B となっ

ていた点 ( $ps < .05$ )<sup>12</sup> が挙げられる。このことより、TPC プログラムの効果が最終時期において確かに現れており、行動レベルの高さという点においては、介入群 A のほうがより高かったことが明らかになった。

## 総合考察

本研究の目的は、高校生を対象としたチームによる課題解決スキルの向上を目的とした TPC プログラムについて研修評価アプローチを用いた効果測定を行うことであった。そのために、Kirkpatrick & Kirkpatrick (2006) の FLA のうちの3段階（反応・学習・行動レベル）に沿った測定指標を開発したうえで、介入群（A・B）と比較群からのデータ収集を行い、プログラムのプロセスとアウトカムに関する効果測定を行った。その結果、TPC プログラムは受講生であった高校生に対して概ね、肯定的な効果を生み出していたことが明らかになった。つまり、当該プログラムは確かにチームによる課題解決のためのチームワークの向上に一定の役割を果たしていたことが実証された。

実施回ごとに若干の違いはあったものの、全般的に FLA における反応レベルが高かったことより、TPC プログラムの内容や実施方法に満足していたことが明らかとなった。また、介入群 A

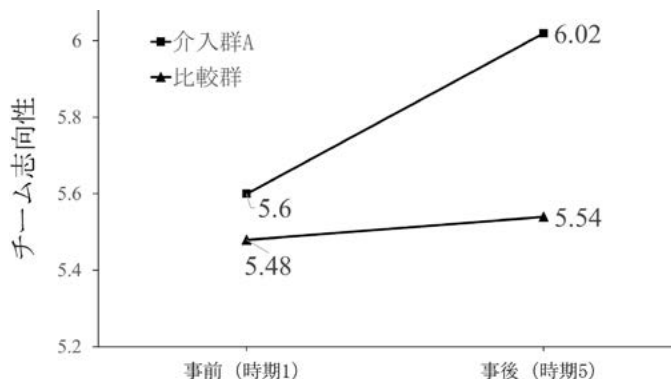


図3 行動レベル（チーム志向性）の変化：不等価（介入群・比較群）事前事後テストデザインによる検討

は介入群 B よりも反応レベルが少し高かったことも確認された。介入群 A は女子高校で介入群 B は共学高校であったことや学びの質や学習能力の違いも何らかの影響を与えていたとも想定される。しかし、その差は効果量にして 15% 程度 ( $\eta^2=.15$ ) に留まっていた。

一方、介入群 A と比較群（同一高校）を比較した分析から、介入群 A はプログラム開始時には満足度が高いが、逆に、終了時には満足度が低くなってしまおうという効果（比較群はその逆のパターン）も認められた。その効果は 11% ( $\eta^2=.11$ ) であった。以上のような留意点があるものの、介入群 A・B ともに各回の満足度が高かったことより、TPC プログラムは全般的に高校生にしっかりと受けとめられていたと考えられる。

先述のとおり、研修プログラムの価値や意義は、満足度（反応レベル）によってのみ評価されるべきものではない、という考えに立脚するのが FLA である。体系的な研修評価には、何を学んだか（学習レベル）とどのような変化が生じたか（行動レベル）という 2 つのレベルにおける効果測定も併せて行うことが必要となる。本研究では、客観（確認）テストによって学習レベルの測定を行ったが、その結果、一部の時期（時期 3）や群・時期（介入群 B における時期 4 と時期 5）において若干低めの得点推移が認められたものの、全体的な理解度は 7 割から 9 割と十分であった。このことより、TPC プログラムの難易度は適切に設定されていたと考えられる。

次に行動レベルの変化であるが、これはプログラム実施にあたってのゴールであったチーム志向性が実際に向上したか否かの評価に関連するものである。本研究の結果、TPC プログラムは総じて高校生のチーム志向性の向上に効果があったことが確認された。受講生の多くにはチーム志向性の向上が認められ、その傾向はプログラムの内容や方法に満足していた人ほど高かったという結果となっていた。一方、各論として、チーム志向性の向上が限定的であったり、予想に反して、逆効果をもたらしたケースも存在した。

特に、プログラムの初期段階でそもそもチーム志向性が高かった介入群 A の高校生にとっては、プログラムの実施がかえってチーム志向性の向上の妨げとなってしまうていたことも明らかになった。これは介入群 A におけるチーム志向性のそもそもの高さ（ie., 切片）と変化（ie., 傾き）の逆相関（表 5 参照）や同群と比較群における講義内容への満足度の変化のパターン（図 1 参照）にも呼応するものである。

したがって、TPC プログラムの改善・向上に際しては、これらのいわゆる“学びの上級者”に対する配慮が必要となってくる。受講者がプログラムに何を求めるのか、あるいは逆に、プログラムが受講者に何を期待するのかを明らかにするために、ニーズの把握やそれに伴ったプログラムのゴール設定や対象者のセレクションが有効であると考えられる（例：三好・安田・梅崎・椋田, 2018）。

本研究では、FLA の枠組みに沿った研修評価アプローチの方法論の有効性も併せて確認できた。昨今、多種多様な研修プログラムが実施されるものの、体系的な効果測定が必ずしも行われていない実情がある。その主な理由として、方法論的に厳格である（methodologically rigorous）ことは科学的探究における前提であるものの、それを実際のフィールドで用いるのには様々な制約が伴うことが挙げられる（例：無作為抽出や無作為配置）。これらの技術的な課題（チャレンジ）を乗り越えるべく提案されたのが FLA である。FLA は、その構造（4 レベル）や連続性（反応から結果まで）が直感的に理解しやすいため、場当たり的ではない、体系的な研修評価の実践の推進という意味においては、有意義なアプローチといえる。また各レベルの指標に関しても比較的スムーズな導入が可能である。

例えば、反応レベルの指標は多くの研修に共通しているため（ie., 満足度の測定）、先例を参考に信頼性・妥当性を担保しつつ、作成または既存の指標を活用することができる。学習レベルについては、客観的に確認テストなどによって数

値化が可能で FLA における分析に役立てられるほかにも、受講者の理解度によっては、プログラムの難易度等の調整の判断材料とすることも可能である。さらに、行動レベルに関して、測定指標はプログラムの活動方針やゴールなどを参考にして作成されるため、受講者におけるアウトカムの効果検証に有用であることはもとより、そもそものプログラムの目的や意義を再確認することにつながってくる。これはプログラムの何をどう評価するのかを予め考えながら、つまり“評価思考 (evaluative thinking)”を行いながらのプログラム開発に通底するところである。

最後に研究の限界と今後の展望について、4つの視点から検討する。第1に、本研究では、FLAの4段階のうち、4段階目（最終レベル）である結果 (results) の測定は行わなかった。プログラムの運営上の課題 (ie., フォローアップ) に起因するものであるが、完全な FLA とするためには、結果レベルの測定も併せて行うことが望ましい。本プログラムにおける結果レベルとしては、例えば、受講者のチーム活動の成果 (例：受賞実績) などがあげられる。

第2に、本研究ではプログラムの対象校が女子高校 (介入群 A・比較群) と共学高校 (介入群 B) であったが、性差または学力レベルによる差の分析は含めなかった。したがって、それらが介入効果の外生的な要因として作用する可能性も十分考えられる。本研究の結果、TPC プログラムの実施効果は両校において確認されたものの、介入群 B においてより発展的な効果が確認されたことも事実である。今後、TPC プログラムがどのような対象に何を伝えていくのかによって、その内容や方法も変化することも想定できる。

第3に本研究では、ベースラインとして時期1データを用いたが、これは初回プログラムの実施後に収集したデータである。そのため、厳密な意味でのベースラインとは言えない。事前データ (ベースライン) の収集には、プログラムを実施する以前からフィールドとの連絡・調整が必要となる。また受講者や受講グループにおいては、プ

ログラムが実施される前から、データ収集への協力を仰ぐことになる。本研究では、受講者の負担を考慮して、事前データではなく時期1データを便宜的なベースラインとした経緯がある。

第4に、本研究における行動レベルの測定概念はチーム志向性であり、これをチームワーク (の高低) の指標とした。しかし、PBL 型における教育プログラムでは自律的なチーム形成を目的としたものが多い。そのようなチームには、チーム志向性 (team orientation) の高さはもとより、社会的な課題の解決を模索する際に考慮すべき概念であるチームエンパワメント (team empowerment) の視点が必要となってくる。今後の展望として、批判的意識 (critical awareness) を重要要素に含むエンパワメント概念<sup>13</sup>を参考にしたチームエンパワメントの測定 (Yasuda et al, 2018) およびプログラム開発・評価への応用が望ましいと言える。

#### 注

- 1 *Teams are social entities composed of members with high task interdependency and shared and valued common goals* (Dyer, 1984 as cited in Salas, Cooke, & Rosen, 2008, p.541). その他、企業などの組織におけるチームは、Kozlowski & Bell (2003) によって、[Teams are] “collectives who exist to perform organizationally relevant tasks, share one or more common goals, interact socially, exhibit task interdependencies, maintain and manage boundaries, and are embedded in an organizational context that sets boundaries, constrains the team, and influences exchanges with other units in the broader entity” (p. 334).
- 2 “While all teams are groups, not all groups can be considered teams” (Parker, 1990 as cited in Hardy and Grace, 1997, p. 2).
- 3 [T]eam members’ collective belief that they



have the authority to control their proximal work environment and are responsible for their team's functioning (Mathieu et al., 2006 p.98).

- 4 エンパワメント概念には他に、エンパワーされる対象についてのプロセス (“empowering”) かアウトカム (“empowered”) かの視点が存在する。
- 5 これは学びに関するアウトプット (結果) とアウトカム (成果・効果) の問題でもあり、学習者の年齢が高ければ高いほど (例: 社会人 vs. 学生 vs. 児童・生徒)、アウトプット (例: 満足度) とアウトカム (例: 知識・スキルの習得) の相関も高くなることが予想できる。
- 6 詳しくは、安田 (2011) の第4部「実験デザインに基づく評価」(pp. 191-228); 安田・渡辺 (2008) の第5章「アウトカム評価」(pp. 101-118), 第6章「実験的手法を用いたプログラム評価」(pp. 119-149) を参照。
- 7 結果レベル (Level 4) への関心も高いが、教育・研修プログラムの実施のみで (よほどそのプログラムの“効き目”が強くない限り) 結果レベルの効果が期待できるとは考えにくい。また短期間で効果があがるとされるプログラムであればあるほど、副次的効果・意図しない効果 (いわゆる“副作用”) への配慮も必要となる。一般に、プログラムの効果測定・評価研究は当該プログラムの効果のみに着目されがちであるが、その副次的効果に関する研究も重要な役割を果たすと考えられる (例: Cook, 1975)。
- 8 知識やスキルの習得が研修におけるアウトカム (指標) とされることが多いが、TPC プログラムでは知識・スキルの習得は最終ゴール (行動レベルの変化) へのステップと位置づけられたため、学習レベルを学びのプロセスと位置づけ、プロセス指標として測定した。
- 9 N=104 (介入群 A) および N=108 (比較群) のデータが分析された。
- 10 Bonferroni の多重比較による分析結果。
- 11 プログラムの効果測定は、カークパトリック

(Kirkpatrick & Kirkpatrick, 2006) の反応 (reaction)、学習 (learning)、行動 (behavior)、結果 (results) の4-level アプローチに基づくものであったが、本研究は TPC プログラムによる行動変容の効果測定が最終ゴールであったため、プログラム実施後の長期の結果 (results) レベルの効果測定は行われなかった。

- 12 Bonferroni の多重比較による分析結果。

- 13 エンパワメント (empowerment) は、自尊心・自尊感情 (self-esteem) や自己肯定感 (self-efficacy) とどこが違うのか。端的には、エンパワメントにあり自尊感情・自己肯定感にないのは、個人をとりまく社会的環境に関する意識、すなわち “Critical Awareness” である (Gutierrez, 1994)。エンパワメント概念に独自かつ必要不可欠なのがこの批判的意識 (critical awareness) である。

#### 参考文献

- 相川充・高本真寛・杉森伸吉・古屋真 (2012) 個人のチームワーク能力を測定する尺度の開発と妥当性の検討『社会心理学研究』27 (3) pp. 139-150. 日本社会心理学会
- Beauchamp, M. R., McEwan, D., & Waldhauser, K. J. (2017). Team building: Conceptual, methodological, and applied considerations. *Current opinion in psychology*, 16, 114-117.
- Bruner, M. W., Eys, M. A., Beauchamp, M. R., & Côté, J. (2013). Examining the origins of team building in sport: A citation network and genealogical approach. *Group Dynamics: Theory, Research, and Practice*, 17(1), 30.
- Cook, T. D. (1975). *Sesame street revisited*. Russell Sage Foundation.
- Gutierrez, L. M. (1994). Beyond coping: An empowerment perspective on stressful life events. *J. Soc. & Soc. Welfare*, 21, 201.
- Hardy, C. J., & Crace, R. K. (1997). Foundations of team building: Introduction to the team building primer. *Journal of Applied Sport*

- Psychology*, 9(1), 1-10.
- Kirkman, B. L., & Rosen, B. (1999). Beyond self-management: Antecedents and consequences of team empowerment. *Academy of Management journal*, 42(1), 58-74.
- Kirkpatrick, D., & Kirkpatrick, J. (2006). *Evaluating training programs: The four levels*. Berrett-Koehler Publishers.
- Klein, C., DiazGranados, D., Salas, E., Le, H., Burke, C. S., Lyons, R., & Goodwin, G. F. (2009). Does team building work?. *Small Group Research*, 40(2), 181-222.
- Kozlowski, S. W. J., & Bell, B. S. (2003). Work groups and teams in organizations. In WC Borman, DR Ilgen, & RJ Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, 333-375).
- Littlepage, G. E., Schmidt, G. W., Whisler, E. W., & Frost, A. G. (1995). An input-process-output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology*, 69(5), 877.
- Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of management*, 34(3), 410-476.
- Mathieu, J. E., Gilson, L. L., & Ruddy, T. M. (2006). Empowerment and team effectiveness: An empirical test of an integrated model. *Journal of applied psychology*, 91(1), 97.
- McGrath, J. E. (1964). *Social psychology: A brief introduction*. Holt, Rinehart and Winston.
- マクリーン S., & ハンソン R. (2019) (木全和巳 訳) 『パワーとエンパワメント』 クリエイツかもがわ
- 縄田健悟・山口裕幸・波多野徹・青島未佳 (2015) 企業組織において高業績を導くチーム・プロセスの解明 『心理学研究』 85-13039.
- 三島一郎 (2007) エンパワメント『日本コミュニティ心理学会 (編) コミュニティ心理学ハンドブック』 pp. 70-84. 東京大学出版会
- 三好真人・安田節之・梅崎修・椋田亜砂美 (2018) 高校生を対象とした Problem Based Learning 型プログラムの体験分析：習熟度別サンプリングによる体験の特徴 日本コミュニティ心理学会第 21 回大会ポスター発表
- 椋田亜砂美・梅崎修・安田節之・三好真人 (2018) < 資料紹介 > 産学連携による PBL 型プログラムの実践報告：サイボウズ社による「TPC プログラム」『生涯学習とキャリアデザイン』15(2) pp. 83-98. 法政大学キャリアデザイン学会
- O'Brien, M. (1995). *Who's got the ball? And other nagging questions about team life*. San Francisco: Jossey-Bass.
- Rappaport, J. (1987). Terms of empowerment/exemplars of prevention: Toward a theory for community psychology. *American journal of community psychology*, 15(2), 121-148.
- Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human factors*, 50(3), 540-547.
- Schein, E. H. (1969). *Process consultation: Its role in organization development*.
- Seibert, S. E., Wang, G., & Courtright, S. H. (2011). Antecedents and consequences of psychological and team empowerment in organizations: A meta-analytic review. *Journal of applied psychology*, 96(5), 981.
- Yasuda, T., Umezaki, O., Mukuta, A., & Miyoshi, M. (2018) Evaluating Teamwork in Japanese High School Students: A measurement tool to assess team empowerment. 126th Annual Convention of American Psychological Association (SFO, CA).
- 安田節之 (2011) 『プログラム評価：対人・コミュニティ援助の質を高めるために』 新曜社.

- 安田節之 (2013) プログラム評価研究：臨床心理サービスのアカウントビリティ向上に役立つ視点（特集 対人援助職の必須知識 研究の方法を知る）－（臨床心理学における量的研究の基礎と発展）『臨床心理学』13（3）pp. 337-342. 金剛出版
- 安田節之 (2014) コミュニティ介入の効果を高める組織特性の検討：臨床心理地域援助における評価研究の試みとして『臨床心理学』14（3）pp. 402-411. 金剛出版
- 安田節之 (2018) 産学連携・PBL 型教育プログラムの評価研究：チームエンパワメントの効果検証に向けて 日本心理学会大会発表論文集『日本心理学会第 82 回大会』pp. 1PM-109. 日本心理学会
- 安田節之・渡辺直登 (2008) 『プログラム評価研究の方法』新曜社
- 山口裕幸 (2008) 『チームワークの心理学：よりよい集団づくりをめざして』サイエンス社

## Appendix A

### モニタリング指標（反応レベル）

1. 授業を楽しく受けることができた。
2. 授業の内容を理解できた。
3. 授業の内容は自分が興味を持っていることであった。
4. 授業の内容を実生活（学校生活）と結びつけて考えることができた。
5. 授業に満足した。
6. 授業で新しい知識を得ることができた。
7. 授業の内容についてもっと知りたいと思った。
8. 授業を真剣に聞いた。
9. 授業で新しい気づきが得られた。
10. 演習（グループワーク）では自分の考えや意見を言うことができた。
11. 演習では他のメンバーの考えや意見を尊重することができた。
12. 演習では他のメンバーとのディスカッションで新しい気づきが得られた。

## Appendix B

### チーム志向性尺度

1. チームの目指す方向を意識して行動する。
2. チームが何らかの成果・結果を出せるように、協力するほうだ。
3. 自分の長所がチームメンバーから認められている。
4. チームでの自分の役割は明確だ。
5. チームでは自分の役割を、怠けることなく果たしている。
6. チームでの話し合いのときに、自分の意見をしっかりと言うほうだ。
7. 自分で分からないことがあれば、他のチームメンバーに聞くことができる。
8. 自分が勝手に行動をとって良いか分からないときには、事前に他のチームメンバーに相談する。
9. チームでの話し合いのなかで意見の食い違いがあった場合、自分と異なる意見にも耳を傾けるほうだ。
10. 結果を出すためにチームで「やること」を決められる。

# **Evaluation of a team building program for Japanese high school students based on the Kirkpatrick's 4-Level Approach**

YASUDA Tomoyuki  
UMEZAKI Osamu  
MUKUDA Asami  
MIYOSHI Masato

---

Teamwork plays fundamental roles in groups and organizations alike. It is a teamwork that enables members of the groups to accomplish more than all of the members could do if each are working alone. Of particular importance to both practitioners and researchers involves with how effective teamwork could be achieved and sustained in early stages of school lives. Furthermore, it is essential to assess/evaluate if and how team development intervention actually work in school settings. The main purpose of this study was to examine the effectiveness of a team-building program for high-school youths by partially adapting a measurement framework based on the Kirkpatrick's 4-Level Approach (FLA). Two intervention groups (N=126 and 90, respectively) received the program along with another group (N=118) received other school curricula in order to be served as a control

(comparison) group. Data were collected at five points in time concerning Level 1 (Reaction), Level 2 (Learning), and Level 3 (Behavior) for the two intervention groups and twice in time (pre and post) concerning Level 3 (behavior) for the control group. Both cross-sectional and longitudinal data analyses based on the quasi-experimental design indicated that the early teambuilding intervention indeed worked by revealing the sustainable effects in Reaction (L1) and Learning (L2) as well as the incremental effects in Behavior (L3). Some different patterns of effects were also evidenced, such that the intervention was especially effective for students whose degrees of teamwork were comparatively low at the first stages of the study; yet not very (and some were adversely) effective for those who already possessed a certain level of it at first stages.